LEARNING TO NARROW WEB RESULTS CAN BE THE KEY TO SUCCESSFUL ON-LINE SEARCHING. PART 2. WHERE IS THE FIRE INFORMATION?

by

Nora H. Jason
Building and Fire Research Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899 USA

NIST

Learning to narrow Web results
can be the key to successful on-line searching

by Nora H. Jason
Manager, Fire Research Information
Services, National Institute of Standards
and Technology

Where are you going to find meaningful fire information?

In the last issue of *Fire Findings* (Vol. 7, No. 4), we discussed using FIREDOC, an electronic card catalog (bibliographic fire database) and other electronic catalogs. Now let's look at some ways to find fire information on the Web, a good tool for locating many different types of information. Software called search engines can help you find the information you need among millions of Web sites.

Everyone has a favorite search engine or two, so we'll look at two popular ones and two megasearch engines. We'll show you the results of practice searches for information about electrical fires and pyrolysis of wood.

You're probably familiar with the Alta Vista and Yahoo search engines, and we'll show you how to reduce the number of hits you get there. We'll also review FAST — the biggest search engine — and Northern Light, a close second in size. *Figure 1* describes these search engines and their strengths.

To narrow the results from some search engines, try using some of the same search techniques as in an electronic card catalog. Why look through 3,000 Web pages when software can narrow the results meaningfully?

Making the best use of search engines

In Vol. 7, No. 4, we reviewed the differences between a bibliographic database and a search engine. As a refresher, a search engine is software that builds indexes from Web pages.

When you ask it a question, the software looks for matches in those indexes. If a search engine ranks the hits, the software assigns the ranking and you see hits with the highest rankings first. Alta Vista does not rank hits, but you can sort the results, for example, by keyword.

A search engine usually includes a tips section that tells you how to improve your searches. To get results from a search engine, you should understand the search engine's techniques and the general Boolean concepts to retrieve meaningful answers. To save time and avoid much of the frustration associated with searching, let's review the primary ways to search using Boolean logic.

The Boolean AND (or +) indicates the terms to include in a search. Example: Smoke AND detectors results in hits that have both terms: smoke, detectors from the keyword field or the same sentence or paragraph resulting in the concept: smoke detectors.

Use the Boolean OR to include similar words. Example: investigations OR investigators. Web pages with either word will be selected.

Boolean NOT (- or AND NOT) excludes word(s). Example: investigations AND NOT investigators. Only Web pages with investigations alone will be selected.

Use the Boolean NEAR to include a word that's not too far from the other word, one that's usually in the same sentence. Word order is not important. Example: arson NEAR gasoline.

Parentheses express the concept "nesting," where you keep like terms together to improve the search. Example: (smoke AND detectors) AND (hospitals OR apartments). In this example, all hits should include the terms smoke and detectors in either hospitals or apartments. The information you need dictates how simple or how complex the search question should be.

If you use the Custom Search Folders in Northern Light, you can drill through the Web pages to increase the specificity of the results. Alta Vista, FAST and Yahoo use different techniques. If the results are too broad or cluttered with unexpected results, check the Tips section and go to Advanced Search Techniques. By using the examples, the idea of Boolean searching (ways to increase the precision) will be easy to understand.

A piece of information to remember is that search engines visit the Web daily. All of the searching for this article was done in the fall of 1999, and if you repeat the searches today, you would note different results.

What were the results of the sample searches?

As many users would do in practice, if a large number of hits appeared, only two screens worth of results were viewed (20 hits) to see if they looked "good." A "good" meant at least half of the hits appeared to have information pertinent to the search and were worth viewing. No other evaluation was made because a user may be interested in technical information, opinions of others or reviews of the topic.

The total hits were not considered as most search engines rank the results. In addition, you should always use the available ways of increasing precision in the results. See page 14 for a list of questions to ask yourself when evaluating the contents of a Web page.

Let's look at the two information

searches we conducted and see how different strategies in the four search engines helped us.

Pyrolysis of wood

We assumed there could be a high number of hits, so we used two ways to limit the number: using the period from January 1, 1995 through October 10, 1999, and including only hits in English (this option was not available in FAST). Here is what we found.

Alta Vista

Simple search: pyrolysis of wood

Results: 1,514,565 hits. The first few hits looked good, but we had too many Web pages to view. Advanced search increased the precision of the results.

Advanced search: pyrolysis AND wood (the Boolean AND)

Results: 1,520 hits. The first 20 hits were not good.

Advanced search: (pyrolysis AND wood)

Results: 1,493 hits. The results were marginally better, even though we used the "nesting" concept.

Advanced search: "pyrolysis of wood"

Results: 54 hits. The first 20 looked good and included Web sites for experts and/or technical information.

Note: The quotation marks resulted in a precision search.

FAST

Simple search: pyrolysis of wood

Results: 2,778 hits. The first 20 appeared good and included links to full-text publications. Still too many hits, though.

Advanced search: All of the Words (exact phrase MUST exist in the document): "pyrolysis of wood"

Results: 85 hits. The first 20 were good, including research activities, government publications and journal articles. The quotation marks resulted in a precision search.

Advanced search: Any of the Words (exact phrase MAY exist in the document): "pyrolysis of wood"

Results: 85 hits, same as above.

Advanced search: All of the Words + pyrolysis + wood

Results: 2,792 hits. The number nearly matched that of the simple search.

Note: The quotation marks resulted in a precision search, even though the advanced search Any of the Words resulted in the same hits as All of the Words.

Northern Light

Simple search: pyrolysis of wood

Results: 2,551 hits. The first 20 were good, including research articles, contacts and other URLs. Using the Custom Search Folders to drill down, we selected the folder, Combustion and Flammability. Results: 14 good hits.

Power search: Selected the category Words in Title as one or more books would be a good way to research a topic quickly: pyrolysis of wood

Results: 4 hits that were good.

Yahoo

Simple search: pyrolysis of wood

Results: 1,105 hits. The first 20 were not good as both words rarely appeared together.

General search: "pyrolysis of wood"

Results: 26 hits. The word order wasn't maintained, but there were some good hits.

Advanced search: Intelligent Default — pyrolysis of wood

Results: 1,105 hits, the same as the simple search.

Advanced search: Match on All Words (using the Boolean AND) — pyrolysis AND wood

Results: 1,105 hits, the same as the simple search with the same drawback.

Note: The Web Pages option was used and the results improved.

Electrical fires

Where is the information for learning how to prevent electrical fires, identify people working on electrical fires or investigating electrical fires? Let's see what the four search engines had available.

Alta Vista

Simple search: electrical fires

Results: 1,345 hits. The first 20 were good, including technical resources, consultants and general information.

Advanced search: electrical AND fires,

Figure 1. Search engines and their strengths

Alta Vista
www.altavista.com
More than 125 million full Web pages, more than 10 million newsgroup articles
  Strengths: a pure search engine; can't browse

FAST
www.alltheweb.com
More than 200 million Web pages indexed and is noted as the alternative to Northern Light, Alta Vista and Inktomi
  Strengths: Very quick; indexes a variety of sites

Northern Light
www.northernlight.com
More than 170 million Web pages indexed
  Strengths: Indexes and classifies Web pages and all .gov and .mil sites (government and military); fee-based service for full-text articles

Yahoo
www.yahoo.com
More than 1 million Web pages
  Strengths: Users and staff submit Uniform Resource Locators (URL); focus is on quality; strong in technology, especially computers and popular culture

---

Part II: Learning to narrow Web results ...

January 1, 1995 through October 10, 1999 and only hits in English.

Results: 31,004 hits. The first 20 hits did not appear to be good

Note: Clicked on Sort By and inserted the phrase electrical fires. The results were good.

FAST

Simple search: electrical fires

Results: 33,749 hits. The first 20 were good, including fire investigation reports, government sites and professional organizations.

Advanced search: All of the Words: "electrical fires"

Results: 1,699 hits; first 20 were good.

Advanced search: Any of the Words — "electrical fires"

Results: 1,699 hits, same as above.

Advanced search: All of the Words: + electrical + fires

Results: 33,749, same as simple search.

Northern Light — the ultimate drill downward example.

Simple search: electrical fires

Results: 239,041 hits. The Custom Search Folder, Fire Investigation, was selected.

Results: 173 hits. The first 20 were good and included seminars, investigation reports and contacts.

Power search: electrical fires

Results: 243,731 hits.

We used various software features to further refine the search —

- Limited the subject to technology — Skipped Limited Documents (any type would be acceptable).

 Selected date range: 01/01/1990 through 11/18/1999

Results: 39,851 items.

- The Custom Search Folder, Fire Investigation, was selected.

Results: 184 hits.

- The Arson folder was selected, with 18 hits.

Both final results were good and not too cumbersome to view.

Yahoo

Simple search: electrical fires

Results: 43 hits. The first 20 were good, including expert witnesses and product development.

General search: "electrical fires"

Results: 3 hits, all good.

Advanced search: Exact phrase match — electrical fires

Results: 3 hits, all good.

Advanced search: Match on All Words: electrical fires

Results: 43 hits; first 20 were good.

Note: The option, Web Pages, improved all of the above strategies.

Here are just a few more thoughts to consider before you use these search engines or new ones.

■ Be specific in your searching. If you want information about the movie, *Gone with the Wind,* don't use the search term movie; use the words in the title. The same thought process applies to work-related information.

■ Often a simple search is as good as or better than the advanced search.

■ Drilling down and finding precise hits works better in some search engines than others. For example, the Custom Search Folders in Northern Light may enhance the results for some and perhaps not for other searches. The Web Pages option in Yahoo enhanced the results of the search. Again, know the search engine(s) you use.

■ Remember, too, there are other sources than the Web for information.

■ Here is another word to add to your vocabulary: Metacrawlers. They do not "crawl" the Web to build listings, they send searches to several search engines at once. Dogpile is an example of a metacrawler (**www.dogpile.com**).

Editor's Note: Nora H. Jason has a Master of Library Science degree from State University of New York in Albany. For more information, call her at (301) 975-6862; fax (301) 975-4052; e-mail nora.jason@nist.gov or write: Fire Research Information Services, National Institute of Standards and Technology, 100 Bureau Drive, Building 224, Room A252, Gaithersburg, Maryland, 20899-8644.

Jason's article is a contribution of NIST and is not subject to U.S. copyright regulations unlike the other articles in *Fire Findings.* Also, commercial products or software identified in this article don't imply recommendations or endorsement by the National Institute of Standards and Technology, nor does it imply they are the best available for the purpose. FF

Evaluate sites for reliability as you browse

How reliable is the Web information you're viewing? In this age of instant information, this is a important question. Here are some thoughts about evaluating a Web site.

- ■ Is the source of the information identified?
- ■ Is the Web page dated and a contact person identified?
- ■ Has the Web page been updated recently?
- ■ Is there evidence of quality control?
- ■ Are opinions identified separately from factual information?
- ■ Do you note any biases?